

1/5/1 (Item 1 from file: 351)
DIALOG(R) File 351:Derwent WPI
(c) 2006 Thomson Derwent. All rts. reserv.



013496619 **Image available**
WPI Acc No: 2000-668560/ 200065
XRPX Acc No: N00-495587

Proxy cache server for providing translated homepage to internet user,
analyzes HTML document received from server to extract translation object
which is translated and then transmitted to browser

Patent Assignee: NEC CORP (NIDE)
Number of Countries: 001 Number of Patents: 001
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
JP 2000276431	A	20001006	JP 9982039	A	19990325	200065 B

Priority Applications (No Type Date): JP 9982039 A 19990325

Patent Details:

Patent No	Kind	Lan Pg	Main IPC	Filing Notes
JP 2000276431	A	14	G06F-013/00	

Abstract (Basic): JP 2000276431 A

NOVELTY - A proxy access unit (304) receives the original HTML document (101) which has a uniform resource locator (URL) designated by browser (200) from world wide web server (100). The received document is analyzed to extract a translation object. A machine translation device (309) translates the translation object and generates a translated HTML document, which is then transmitted to browser (200).

USE - For providing translated homepage to internet user.

ADVANTAGE - Performs accurate and quick translation of HTML document for displaying on a browser.

DESCRIPTION OF DRAWING(S) - The figure shows the block diagram of components of browsing system with translator ability.

HTML document (10)

WWW server (100)

Browser (200)

Proxy access unit (304)

Machine translation device (309)

pp; 14 DwgNo 1/14

Title Terms: CACHE; SERVE; TRANSLATION; USER; DOCUMENT; RECEIVE; SERVE;
EXTRACT; TRANSLATION; OBJECT; TRANSLATION; TRANSMIT

Derwent Class: T01

International Patent Class (Main): G06F-013/00

International Patent Class (Additional): G06F-017/28

File Segment: EPI

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号
特開2000-276431
(P2000-276431A)

(43)公開日 平成12年10月6日(2000.10.6)

(51)Int.Cl. ⁷	識別記号	F I	テーマコード [*] (参考)
G 0 6 F 13/00 17/28	3 5 5	G 0 6 F 13/00 15/20 15/38	3 5 5 5 B 0 0 9 5 9 2 A 5 B 0 8 9 Z 5 B 0 9 1

審査請求 有 請求項の数10 O L (全 14 頁)

(21)出願番号 特願平11-82039

(22)出願日 平成11年3月25日(1999.3.25)

(71)出願人 000004237

日本電気株式会社
東京都港区芝五丁目7番1号

(72)発明者 栗原 浩

東京都港区芝五丁目7番1号 日本電気株
式会社内

(74)代理人 100065385

弁理士 山下 稔平

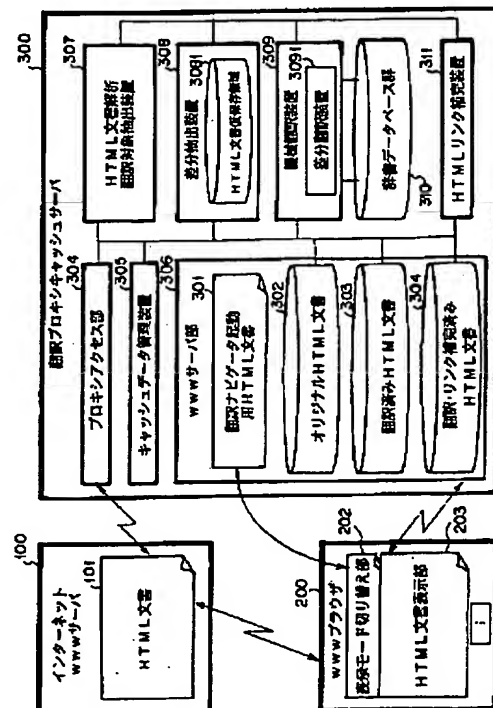
Fターム(参考) 5B009 QA11 QB11 QB18 SA03 SA14
TB14 VB01 VCO1
5B089 GA19 GB04 JB02 KA02 KC53
KC60 KD02 KH03 KH17
5B091 AA03 BA03 CD15 DA02 EA12
EA21

(54)【発明の名称】 翻訳機能付きプロキシキャッシュサーバ及びそれを有する翻訳機能付きブラウジング・システム

(57)【要約】

【課題】 インターネット上に存在する最新のHTML文書を効率良く高速に翻訳し、さらにはインターネット上に接続している全てのクライアントコンピュータに対して、それらが機械翻訳プログラム・辞書データベースを有していない場合でも翻訳済み最新HTML文書をブラウザ上に任意に表示させる。

【解決手段】 翻訳プロキシキャッシュサーバは、ブラウザにより指定されるURLを有する原HTML文書をWWWサーバより受信する手段と、前記原HTML文書を解析して翻訳対象を抽出する手段と、前記翻訳対象を翻訳して翻訳済みHTML文書を生成する翻訳手段と、を備え、前記翻訳済みHTML文書を前記ブラウザに送信する。ブラウザは、サーバに所望のHTML文書のURLを指示する手段を備える。



【特許請求の範囲】

【請求項 1】 ブラウザにより指定される URL (Uniform Resource Locator) を有する原 HTML 文書を WWW (World Wide Web) サーバより受信する手段と、前記原 HTML 文書を解析して翻訳対象を抽出する手段と、前記翻訳対象を翻訳して翻訳済み HTML 文書を生成する翻訳手段と、を備え、前記翻訳済み HTML 文書を前記ブラウザに送信することを特徴とする翻訳機能付きプロキシキャッシュサーバ。

【請求項 2】 請求項 1 に記載の翻訳機能付きプロキシキャッシュサーバにおいて、前記翻訳済み HTML 文書に前記原 HTML 文書又は前記翻訳済み HTML 文書を選択するための補完情報を付加して翻訳・リンク補完済み HTML 文書を生成する手段を更に備え、前記翻訳済み HTML 文書の代わりに、前記翻訳・リンク補完済み HTML 文書を送信することを特徴とする翻訳機能付きプロキシキャッシュサーバ。

【請求項 3】 請求項 1 又は 2 に記載の翻訳機能付きプロキシキャッシュサーバにおいて、前記原 HTML 文書を保持する原 HTML 文書保持手段と、WWW サーバから受信した前記原 HTML 文書と、該 WWW サーバから受信した原 HTML 文書と同一の URL を有する前記 HTML 文書保持手段に保持されている前記原 HTML 文書との差分を抽出する手段と、を更に備え、前記翻訳手段は、前記差分のみを翻訳し、翻訳結果を前記翻訳済み HTML 文書に併合させ、前記原 HTML 文書保持手段は、新たに WWW サーバから受信した前記原 HTML 文書を保持することを特徴とする翻訳機能付きプロキシキャッシュサーバ。

【請求項 4】 請求項 3 に記載の翻訳機能付きプロキシサーバにおいて、前記 WWW サーバ上の原 HTML 文書の最終更新日と、前記 WWW サーバ上の原 HTML 文書と同一の URL の前記原 HTML 文書保持手段に保持されている原 HTML 文書の最終更新日とを比較する手段を更に備え、前記 WWW サーバ上の原 HTML 文書の最終更新日が、原 HTML 文書保持手段に保持されている原 HTML 文書の最終更新日と同一であるときには、前記翻訳手段は翻訳済み HTML 文書の生成または、前記併合による更新をおこなわないことを特徴とする翻訳機能付きプロキシキャッシュサーバ。

【請求項 5】 請求項 1 乃至 4 のいずれか 1 項に記載の翻訳機能付きプロキシキャッシュサーバにおいて、前記 URL が相対指定である場合に、当該 URL を絶対指定に変換する手段を更に備えることを特徴とする翻訳機能付きプロキシキャッシュサーバ。

【請求項 6】 請求項 1 乃至 5 に記載の翻訳機能付きプロキシキャッシュサーバと、該サーバに所望の HTML 文書の URL を指示する手段を備えるブラウザと、を備えることを特徴とする翻訳機能付きブラウジング・システム。

【請求項 7】 請求項 2 乃至 5 に記載の翻訳機能付きプロキシキャッシュサーバと、該サーバに所望の HTML 文書の URL を指示する手段と、原 HTML 文書又は翻訳済み HTML 文書を指定する手段と、を備えるブラウザと、を備えることを特徴とする翻訳機能付きブラウジング・システム。

【請求項 8】 請求項 6 又は 7 に記載の翻訳機能付きブラウジング・システムにおいて、前記ブラウザ上で翻訳ナビゲータ起動用 HTML 文書により実現されるナビゲータが動作し、該ナビゲータが前記 URL を前記サーバに指示することを特徴とする翻訳機能付きブラウジング・システム。

【請求項 9】 請求項 8 に記載の翻訳機能付きブラウジング・システムにおいて、前記翻訳機能付きプロキシキャッシュサーバが翻訳ナビゲータ起動用 HTML 文書を保持し、前記ブラウザは、前記翻訳機能付きプロキシキャッシュサーバより前記翻訳ナビゲータ起動用 HTML 文書を受信することを特徴とする請求項 8 に記載の翻訳機能付きブラウジング・システム。

【請求項 10】 コンピュータを請求項 1 乃至 5 のいずれか 1 項に記載の翻訳機能付きプロキシキャッシュサーバとして機能させるためのプログラムを記録したことを特徴とするコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、ホームページをブラウジングするためのブラウジング・システムに関し、特に翻訳されたホームページをブラウジングするためのブラウジング・システム及びそのためのプロキシキャッシュサーバに関する。

【0002】

【従来の技術】現在インターネット上にはあらゆる情報を発信する WWW (World Wide Web) サーバが世界中に存在し、その情報はインターネットに接続したクライアントコンピュータの WWW ブラウザ (以下ブラウザ) を使って容易に取得することができる。

【0003】インターネット上の情報は一般的に HTML (HyperText Markup Language) で記述された文章と、画像ファイルによって構成されており、これらはホームページ、WEB ページと呼ばれている。

【0004】これらインターネット上の情報の所在は URL (Uniform Resource Locator) という記述方式によって一意に決定される。URL によって指定され、ブラウザ上に表示された HTML 文書内の文字列はさらに「リ

ンク」という形でURLが関連付けられ、ブラウザを操作しているユーザーはこのリンクされた文字列をクリックすることで、更なる情報をブラウザ上に表示させることが可能である。

【0005】このようなインターネット上の情報を、リンクをクリックしながら取得していく操作はブラウジング、ネットサーフィンなどと呼ばれている。

【0006】インターネット上の情報の殆どは英語で記述されたHTML文書であり、この表示された英文を英語以外の言語（つまり母国語）に翻訳して表示させる為の技術、ソフトウェアが注目されている。

【0007】一般的にWEBページ翻訳ソフトウェアと呼ばれるプログラムの実施形態の1つは、ブラウジングを行っているユーザーのコンピュータに翻訳ソフトをインストールして、この翻訳ソフトウェアとブラウザとを連動させて表示中の英語で描かれた文章を翻訳して表示するというものである。このようなソフトウェアとしては、(株)富士通の「翻訳サーフィン」、(株)沖ソフトウェアの「PENSEE for Internet」、(株)日立製作所「たちまち翻訳」等が市販されている。この形態のソフトウェアの中には、インターネットに接続していなくても、各クライアントのハードディスクに保存（キャッシュ）された既に取得済みのWEBページを翻訳して表示できるものもあり、この技術は特開平9-146818号公報に記載されている。

【0008】

【発明が解決しようとする課題】この形態のソフトウェアの第1の問題点は、ブラウザを利用するクライアントコンピュータの1つ1つに翻訳エンジンを有するプログラムと、そのプログラムが利用する辞書データベースを保持していなくてはならないということである。

【0009】例えば多くのコンピュータを有する企業においてこの形態のWEBページ翻訳ソフトウェアの各コンピュータ上での利用は多くのコストと多くのメンテナンス作業を発生させ、非現実的である。

【0010】第2の問題点は、各クライアントコンピュータにインストールされた翻訳ソフトウェアの翻訳の精度が悪いということである。翻訳の精度を上げるには、利用する辞書の種類・規模が翻訳対象である文章の分野に関連されている必要があると一般的に言われている。

【0011】しかし1つ1つのクライアントのマシンに膨大な種類・量の辞書と、翻訳対象と辞書の種類を学習させて保持する仕組みを持たせることは、非現実的である。

【0012】これに対し、翻訳装置をクライアントコンピュータに持たずに翻訳の精度を向上させる技術が特開平10-134052号公報に記載されている。これはブラウザのプロキシサーバの機能を利用した翻訳システムである。

【0013】ブラウザのプロキシサーバの設定を翻訳ブ

ロキシサーバである機械翻訳装置を有するWWWサーバに指定し、ユーザーの指定したURLのHTML文書をプロキシサーバが代理となって取得し、取得した文書を翻訳してブラウザに表示させるというものである。さらに指定されたURLとユーザーの指定した辞書とを関連付けて、その後の翻訳には対応する辞書を利用することで精度を上げるというものである。

【0014】この形態の翻訳システムは、クライアントの環境（ここでいう環境とは、ブラウザを起動しているコンピュータの性能、翻訳ソフトウェア、辞書の有無などを指す。）に依存せず、精度の高い翻訳済みHTML文書をブラウザ上に表示させることが可能であるが、プロキシサーバが一方的に翻訳を行ってしまうという問題がある。プロキシサーバの設定は一般的にブラウザの設定で行うものであり、この設定はブラウザに表示されているHTML文書毎に変更するものではなく、設定している限り全てこのプロキシサーバを経由したブラウジングを行うことになる。これは表示させるHTML文書の言語に関係なく一方的に翻訳処理が行われてしまう問題となる。

【0015】更に、この形態の翻訳システムにおいては、プロキシサーバでの翻訳処理に多くの時間が費やされ、HTML文書を表示するまでにかなりの時間を要してしまう問題がある。

【0016】当然プロキシサーバを設定している限り翻訳を必用としないHTML文書を表示するのに要する時間も増大してしまうことになる。

【0017】更に、この形態の翻訳システムにおいては、プロキシサーバを指定している限り、実際のURLで指定したオリジナルHTML文書（原HTML文書）はブラウザ上に一切表示されないこととなり、真実の情報を覆してしまうことになりかねない。

【0018】本発明は、上記問題を解決し、インターネット上に存在する最新のHTML文書を効率良く高速に翻訳し、さらにはインターネット上に接続している全てのクライアントコンピュータに対して、それらが機械翻訳プログラム・辞書データベースを有していない場合でも翻訳済み最新HTML文書をブラウザ上に任意に表示させることが可能な翻訳機能付きブラウジング・システムを提供することを目的とする。

【0019】

【課題を解決するための手段】本発明による翻訳機能付きプロキシキャッシュサーバは、ブラウザにより指定されるURL(Uniform Resource Locator)を有する原HTML文書をWWW(World Wide Web)サーバより受信する手段と、前記原HTML文書を解析して翻訳対象を抽出する手段と、前記翻訳対象を翻訳して翻訳済みHTML文書生成する翻訳手段と、を備え、前記翻訳済みHTML文書を前記ブラウザに送信することを特徴とする。

【0020】また、本発明による翻訳機能付きプロキシ

5

キャッシュサーバは、上記の翻訳機能付きプロキシキャッシュサーバにおいて、前記翻訳済みHTML文書に前記原HTML文書又は前記翻訳済みHTML文書を選択するための補完情報を付加して翻訳・リンク補完済みHTML文書を生成する手段を更に備え、前記翻訳済みHTML文書の代わりに、前記翻訳・リンク補完済みHTML文書を送信することを特徴とする。

【0021】更に、本発明による翻訳機能付きプロキシキャッシュサーバは、上記の翻訳機能付きプロキシキャッシュサーバにおいて、前記原HTML文書を保持する原HTML文書保持手段と、WWWサーバから受信した前記原HTML文書と、該WWWサーバから受信した原HTML文書と同一のURLを有する前記HTML文書保持手段に保持されている前記原HTML文書との差分を抽出する手段と、を更に備え、前記翻訳手段は、前記差分のみを翻訳し、翻訳結果を前記翻訳済みHTML文書に併合させ、前記原HTML文書保持手段は、新たにWWWサーバから受信した前記原HTML文書を保持することを特徴とする。

【0022】更に、本発明による翻訳機能付きプロキシキャッシュサーバは、上記の翻訳機能付きプロキシキャッシュサーバの翻訳機能付きプロキシサーバにおいて、前記WWWサーバ上の原HTML文書の最終更新日と、前記WWWサーバ上の原HTML文書と同一のURLの前記原HTML文書保持手段に保持されている原HTML文書の最終更新日とを比較する手段を更に備え、前記WWWサーバ上の原HTML文書の最終更新日が、原HTML文書保持手段に保持されている原HTML文書の最終更新日と同一であるときには、前記翻訳手段は翻訳済みHTML文書の生成または、前記併合による更新をおこなわないことを特徴とする。

【0023】本発明による翻訳機能付きプロキシキャッシュサーバは、上記の翻訳機能付きプロキシキャッシュサーバにおいて、前記URLが相対指定である場合に、当該URLを絶対指定に変換する手段を更に備えることを特徴とする。

【0024】本発明による翻訳機能付きブラウジング・システムは、上記の翻訳機能付きプロキシキャッシュサーバと、該サーバに所望のHTML文書のURLを指示する手段を備えるブラウザと、を備えることを特徴とする。

【0025】また、本発明による翻訳機能付きブラウジング・システムは、上記の翻訳機能付きプロキシキャッシュサーバと、該サーバに所望のHTML文書のURLを指示する手段と、原HTML文書又は翻訳済みHTML文書を指定する手段と、を備えるブラウザと、を備えることを特徴とする。

【0026】更に、本発明による翻訳機能付きブラウジング・システムは、上記の翻訳機能付きブラウジング・システムにおいて、前記ブラウザ上で翻訳ナビゲータ起

6

動用HTML文書により実現されるナビゲータが動作し、該ナビゲータが前記URLを前記サーバに指示することを特徴とする。

【0027】更に、本発明による翻訳機能付きブラウジング・システムは、上記の翻訳機能付きブラウジング・システムにおいて、前記翻訳機能付きプロキシキャッシュサーバが翻訳ナビゲータ起動用HTML文書を保持し、前記ブラウザは、前記翻訳機能付きプロキシキャッシュサーバより前記翻訳ナビゲータ起動用HTML文書を受信することを特徴とする。

【0028】本発明によるコンピュータ読み取り可能な記録媒体は、コンピュータを上記の翻訳機能付きプロキシキャッシュサーバとして機能させるためのプログラムを記録したことを特徴とする。

【0029】

【発明の実施の形態】本発明は、インターネット上に存在するHTML文書を翻訳してブラウザに表示する手段を提供するものであり、日々更新されるHTML文書を効率的に翻訳し、インターネット上に接続している全てのクライアントコンピュータに同様の翻訳済みHTML文書を高速に提供する構成を提供するものである。

【0030】図1において、インターネット上に存在するHTML文書01はWWWサーバ1のディスク装置に保存され、WWWサーバによって発信されている。HTML文書101はインターネットに接続されたクライアントコンピュータのWWWブラウザ（以下、「ブラウザ」という。）200にダウンロードされ表示される。

【0031】このHTML文書を翻訳してブラウザ表示させる為に、翻訳プロキシキャッシュサーバ（以下、「翻訳サーバ」という。）300にアクセスし、翻訳文書とオリジナル文書とをブラウザ200上で切り替えるための翻訳ナビゲータ用HTML文書101をダウンロードし、表示モード切り替え部202に表示させると共に翻訳対象HTML文書のURLを翻訳サーバ300に送信する。

【0032】翻訳サーバ300では、キャッシュデータ管理装置305が指定されたURLのHTML文書101が翻訳サーバ300のディスク装置にオリジナルHTML文書302として保存されているかどうかをチェックした後、プロキシアクセス部304が指定されたURLのHTML文書101にアクセスする。

【0033】翻訳サーバ300内では、キャッシュデータ管理装置305と指定されたURLのHTML文書101の最新情報を基に、以下の3パターンの処理が行われる。

【0034】1. [指定されたHTML文書101がオリジナルHTML文書キャッシュ302として保存されていない場合] プロキシアクセス部304はHTML文書101のダウンロードを行い、オリジナルHTML文書302として保存する。

【0035】ダウンロードされたHTML文書101からは、HTML文書解析・翻訳対象抽出装置307によって解析され翻訳対象である文書が抽出される。

【0036】辞書データ・データベースを有する機械翻訳装置309は、抽出された翻訳対象文書の翻訳を行い、オリジナルHTML文書の構成情報を記述するタグを付与した後、WWWサーバ300のディスク装置に翻訳済みHTML文書303として保存する（翻訳処理）。

【0037】さらにHTMLリンク補完装置311では、翻訳済みHTML文書のハイパーテキスト用のタグ記述（以下リンク）に対し、ブラウザ200上の表示モード切り替え部202が「翻訳モード」「オリジナルモード」の切り替えが可能なような記述を補完し、WWWサーバ300のディスク装置に翻訳・リンク補完済みHTML文書303として保存する（HTMLリンク補完処理）。

【0038】2. [指定されたHTML文書01がオリジナルHTML文書キャッシュ002に保存されている文書と異なる場合] プロキシアクセス部304はHTML文書101のダウンロードを行い、差分抽出装置308がダウンロードしたHTML文書101とオリジナルHTML文書302との差分を抽出した後、HTML文書101をオリジナルHTML文書002として保存する。

【0039】抽出した差分情報を元に、機械翻訳装置309の差分翻訳装置006が変更のあった文書構成に対してのみ翻訳処理を行い、前回翻訳された翻訳済みHTML文書に対して変更部分をマージし、翻訳済みHTML文書303として保存する（差分翻訳処理）。

【0040】HTMLリンク補完装置311はHTMLリンク補完処理を行い、WWWサーバ300のディスク装置に翻訳・リンク補完済みHTML文書304として保存すると共に、ブラウザ200に送信する。

【0041】3. [指定されたHTML文書01がオリジナルHTML文書キャッシュ002に保存されている文書と同一である場合] プロキシアクセス部304は指定されたHTML文書01に対応する翻訳・リンク補完済み文書304をブラウザ200に送信する。

【0042】以上が、それぞれの場合の翻訳プロキシキャッシュサーバの動作である。

【0043】ブラウザ200上では指定されたHTML文書の翻訳・リンク補完済み文書3004を受信し、HTML文書表示部203に翻訳された文書を表示する。

【0044】HTML文書表示部203に表示されている翻訳・リンク補完済みHTML文書内リンクが選択されると、表示モード切り替え部202の翻訳ナビゲータが翻訳モードかオリジナルモードかを判別し、翻訳モードの場合は翻訳サーバ300に対して選択したURLを送信する。オリジナルモードの場合は通常のブラウジン

グと同様に選択したURLに対してアクセスする。

【0045】このようにして、インターネット上のHTML文書の翻訳を、クライアントに翻訳ソフトを持たずして、かつ高速に表示させることを可能とする。

【0046】[実施形態1] 図1において、本発明の構成は、インターネット上のWWWサーバ100に存在するHTML文書101のURLを指定して表示させるブラウザ200と、指定されたURLに対して代理にアクセスし、指定されたHTML文書の翻訳された文書を送信する翻訳サーバ300に大別される。

【0047】ブラウザ200は、CPU、ディスク装置及びメモリを有するコンピュータ上で動作し、URLを指定しHTTPプロトコルを使用してデータをダウンロードしてHTMLを解析して表示する機能を有し、ブラウザ上に表示中のHTML文書のURLの他、このブラウザの実行中に訪れたURLの履歴をコンピュータ上のディスク及びメモリ上に保持することのできる一般的なWWWブラウザである。

【0048】さらにブラウザ200は、HTML文書に記述されたプログラム言語について解釈・コンパイルを実行し、ダウンロード後に例えば表示するHTML文書のURLの変更やブラウジングの履歴の取得や別の表示ウィンドウの起動などの機能を実行可能なWWWブラウザである。

【0049】さらにこのクライアントコンピュータにはキーボード等のURLを入力するための入力装置、マウス等のブラウザやブラウザ上のHTML文書のリンクを操作する為のポインティング装置、及びブラウザのユーザーインタフェースを表示させる為のディスプレイ装置に接続されているものである。

【0050】翻訳サーバ300は、CPU、ディスク装置及びメモリを有するコンピュータであり、指定されたURLに接続してダウンロードを行うプロキシアクセス部304と、クライアントコンピュータ上のブラウザ200に対して翻訳ナビゲータ起動用HTML文書301や、翻訳・リンク補完済みHTML文書を送信するWWWサーバ部306と、ディスク装置にキャッシュとして保存されているHTML文書の更新日とそのURLを管理するキャッシュデータ管理装置305と、HTML文書を保存する為のディスク装置、すなわち保存される文書によって分ければ翻訳前のオリジナルHTML文書302を保存するディスク装置と、オリジナルHTML文書302の構成情報とその翻訳文書である翻訳済みHTML文書303を保存するディスク装置と、翻訳済みHTML文書内のリンクをブラウザ200の表示モード切り替え部における翻訳ナビゲータが解釈可能なリンクタグに変換した翻訳・リンク補完済みHTML文書304を保存するディスク装置、ダウンロードされたHTML文書と、翻訳サーバ内で保存されているオリジナル文書の差分情報を抽出する差分抽出装置308と、ダウンロ

ードされたHTML文書を解析し、その段落構成やリンク等の構成を判別すると共に機械翻訳装置に対して翻訳処理の対象となる文書を抽出するHTML文書解析翻訳対象抽出装置307と、機械翻訳用の各種辞書データをディスクに保持する辞書データベース群310を有し、翻訳対象HTML文書の構成を示すタグと機械翻訳された文書で翻訳済みHTML文書303を生成する機械翻訳装置309と、差分翻訳装置306において、オリジナルHTML文書の差分情報を元に、修正されている部分のみを翻訳し、翻訳済みHTML文書003にマージ

(併合)して最新の翻訳済みHTML文書を生成する差分翻訳装置306と、翻訳済みHTML文書303のハイパーテキストリンクに対して、ブラウザ上の表示モード切り替え部が「翻訳モード」「オリジナルモード」の切り替えが可能なような記述を補完する機能を有するHTMLリンク補完装置311で構成される。

【0051】次に、翻訳サーバ300上の本実施形態の動作について、図1、図2を参照して説明する。

【0052】翻訳サーバ300は、翻訳対象となるHTML文書のURL文字列を取得する(S001)。

【0053】キャッシュデータ管理装置305は、指定されたURLのHTML文書101がオリジナルHTML文書302として保存されているかどうかをチェックする。(S002) S002においてキャッシュされている文書であると判断された場合、プロキシアクセス部304が指定されたURLのHTML文書01にアクセスし、最終更新日を取得する(S003)。

【0054】キャッシュデータ管理装置305は、指定されたURLのHTML文書01の最終更新日がオリジナルHTML文書002として保存されているキャッシュデータの最終更新日と同一であるかどうかチェックする(S004)。

【0055】S004においてキャッシュデータが最新と判断された場合、WWWサーバ部306が、対応する翻訳・リンク補完済HTML文書304をディスク装置より取得し(S005)、要求のあったブラウザへ送信する(S006)。

【0056】S002においてキャッシュされていない文書であると判断された場合、プロキシアクセス部304が指定されたURLのHTML文書101にアクセスし、最終更新日を取得し(S017)、キャッシュデータ管理装置305のデータとしてURLと最終更新日が保存される(S018)。

【0057】プロキシアクセス部304は、指定されたURLのHTML文書のダウンロードを行い(S019)、ダウンロードしたオリジナルHTML文書は、ディスク装置にオリジナルHTML文書302として保存される(S020)。

【0058】次に、ダウンロードされたオリジナルHTML文書に対して、翻訳処理が行われる(S021)。

翻訳処理については後述する。

【0059】S021において翻訳されたHTML文書は、ディスク装置に翻訳済みHTML文書303として保存される(S014)。

【0060】翻訳済みHTML文書内のハイパーテキストリンクは、HTMLリンク補完装置311によってブラウザ上の表示モード切り替え部が「翻訳モード」「オリジナルモード」の切り替えが可能なような記述へと変更される(S015)。リンク補完処理については後述する。

【0061】S015においてリンク補完処理が行われたHTML文書は、ディスク装置に翻訳・リンク補完済みHTML文書304として保存され(S016)、WWWサーバ部306が要求のあったブラウザへ送信する(S006)。

【0062】S004においてキャッシュデータが最新ではないと判断された場合、ディスクに保存されているオリジナルHTML文書302を翻訳サーバ300のHTML文書仮保存領域3081に移動する(S007)。

【0063】プロキシアクセス部304は、指定されたURLのHTML文書01にアクセスし、最終更新日を取得し(S008)、キャッシュデータ管理装置のデータとしてURLと最終更新日を保存し(S009)、指定されたURLのHTML文書のダウンロードを行う(S010)。

【0064】S010でダウンロードした最新のオリジナルHTML文書は、S007でHTML文書仮保存領域3081に移動した前回ダウンロードした時のオリジナルHTML文書との差分情報が差分抽出装置308によって抽出(S011)された後、オリジナルHTML文書302として新たに保存される(S012)。同時に差分抽出用にHTML文書仮保存領域3081に移動していた前回ダウンロードした時のオリジナルHTML文書は削除される。

【0065】S011で生成された差分情報を元に、変更のあった文書構成のみを翻訳し、翻訳済みHTML文書303にマージ(併合)する差分翻訳処理を行う(S013)。差分翻訳処理については後述する。

【0066】S013において変更部分が翻訳されマージ(併合)されたHTML文書は、ディスク装置に翻訳済みHTML文書303として保存される(S014)。

【0067】S015においてリンク処理が行われたHTML文書は、ディスク装置に翻訳・リンク補完済みHTML文書304として保存され(S016)、WWWサーバ部306が要求のあったブラウザへ送信する(S006)。

【0068】[翻訳処理] HTMLの文法では図3に示すように「<」と「>」に囲まれている部分がタグとな

っており、この内部の記述はブラウザに表示される本文の内容には無関係である。よってこれを文章として翻訳するにはこのタグを除去した文章を生成し、翻訳の後に再度タグを付与できるようなコメントを翻訳対象文書に付加する必要がある、これを行うのがHTML文書解析・翻訳対象抽出装置307である。図4にHTMLタグの解析と翻訳後のHTMLタグの復活の流れを示す。

【0069】オリジナルHTML文書 文1のタグ内の文章を取り出し、「/A」のタグを除去したものが文2である。文2のタグ内の文章を保持している部分と、文章を翻訳したものが文3である。タグ内の文章を実際のリンクタグへ戻したものが文4の翻訳済み文書である。

【0070】さらに、差分翻訳処理において変更のあった文、もしくは段落などの文章の構成単位のみ翻訳を行う為には、オリジナルHTML文書と翻訳済みHTML文書の構成の対応をとる為の構成情報を翻訳済みHTML文書に付与する必要がある。HTMLの文法では「<!--」と「>」で囲まれている部分がコメントとされ、ブラウザでは無視される部分である。このコメントタグを利用して文章構成情報を付与した例を図5に示す。この例の場合段落を現すHTMLのタグ「<P>

」を一つの文書構成と位置づけ、コメントタグに文書構成名とその文書構成のスタートの位置としてオリジナルHTML文書における行数と桁数を記述している。この文書構成と翻訳前の差分情報により差分翻訳処理を行うことが可能である。

【0071】なお、文書構成の単位としては、「空行」や改行を現す「
」等HTMLの文法上様々なタグによって分けることが可能であり、よって様々な文書構成単位が考えられる。

【0072】[差分翻訳処理] 差分抽出装置308によって抽出された差分情報の例を図6に示す。これはコンピュータワークステーションの標準的なオペレーティングシステムであるUNIXの一般的なコマンドである「diff」コマンドによる2つのファイルの差分の表示結果であり、例では12行目の変更が示されている。なお差分情報の抽出は、変更がかかった部分の2つのファイルにおける変更位置と変更内容が取得できれば方法は限定しない。

【0073】図6の抽出された差分情報によれば、オリジナルHTML文書の12行目の内容が変更されている。ここで差分翻訳装置006は、変更前のオリジナルHTML文書の12行目が翻訳済みHTML文書のどの文書構成に属する文章かどうかを取得し、変更のあった文書構成のみを翻訳対象HTML文書より抽出し翻訳を行う。この翻訳済み文書を文書構成単位毎に既にある翻訳済みHTML文書にマージ(併合)することで差分翻訳処理は行われる。

【0074】[HTMLリンク補完処理] 翻訳処理S021、及び差分翻訳処理S013によって生成された翻

訳済みHTML文書のハイパーテキスト用のHTMLタグ記述(以下、「リンク」という。)は、そのままではブラウザ2上でこのリンクをクリックしても本来アクセスすべきHTML文書を表示させることはできないタグの記述となっている場合がある。

【0075】図7を用いて説明する。リンク1はドメイン名まで記述したリンクであり、これをブラウザ上でクリックすると指定されたURLに接続しHTML文書をダウンロード可能であるが、これは通常のブラウジングであり、このHTML文書を翻訳させるには再度表示中のHTML文書のURLを翻訳サーバに送信する必要がある。

【0076】リンク2、3は表示中のHTML文書と相対的なHTML文書のパスを指定しており、これをブラウザ上でクリックすると、その文書が存在していればその文書が表示されるが、存在しない場合はエラーとなってしまう。存在した場合でもそれは実際インターネット上で公開されている最新のHTML文書ではなく、単にキャッシュとして保存されている文書にすぎない。

【0077】HTMLリンク補完装置11は、このようなリンクにドメイン名を補完し、さらにブラウザ上の表示モード切り替え部が「翻訳モード」「オリジナルモード」の切り替えが可能なようなタグへ変更を行う。変更されたタグの例を図8に示す。対象となるHTML文書のオリジナルHTML文書のURLが「http://www.???.co.jp/AAA/BBB/sample.html」である場合、各リンクのURLは「http://...」で始まるドメイン名の入ったURLに変換される。さらに図8における「onClick = function(this)」の記述は、このリンクがブラウザ上で選択された時にfunction()という関数にこのリンクのURLを引数として渡していることを示すものである。このfunction()は翻訳ナビゲータのfunction()であると想定しており、「翻訳モード」「オリジナルモード」の判別を行い、渡されたURLを翻訳サーバに送信するかURLに対して直接接続する翻訳ナビゲータの機能を実行するものである。

【0078】次に、WWWブラウザ200上の本実施形態の動作について、図1、図9、図10を用いて説明する。

【0079】まずブラウザ200上に翻訳を行いたいHTML文書(翻訳対象HTML文書)を表示させる(S101)。

【0080】次にブラウザ200上で翻訳サーバ300の翻訳ナビゲータ起動用HTML文書001のURLを指定し、翻訳サーバ3へアクセスする(S102)。

【0081】翻訳サーバ300は翻訳ナビゲータ起動用HTML文書301を送信する(S103)。

【0082】ブラウザ200上に翻訳ナビゲータ起動用HTML文書301がダウンロードされる。この時ブラウザ200には翻訳ナビゲータHTML文書301が表

示されている表示モード切り替え部 202 と、翻訳ナビゲータの管理可能な HTML 文書表示部 203 の 2 つ以上の HTML 文書表示部が表示されている。これを総称して翻訳モードでの翻訳ナビゲータの起動とする (S104)。

【0083】翻訳ナビゲータは、起動と同時にブラウザ 200 上の履歴をたどり、1 つ前に表示していた HTML 文書、つまり S001 で読み込んだ翻訳対象 HTML 文書の URL を取得し翻訳サーバに送信する (S105)。

【0084】さらに翻訳ナビゲータは翻訳対象 HTML 文書を HTML 文書表示部に表示させる (S106)。

【0085】翻訳サーバ 300 では翻訳対象 HTML 文書の URL を取得し、翻訳処理が行われた後、翻訳・リンク補完済み HTML 文書 304 を送信する (S107)。

【0086】ブラウザ 200 は翻訳済み・リンク補完済み HTML 文書 304 を HTML 文書表示部 203 に表示する (S108)。

【0087】さらに表示された翻訳・リンク補完済み HTML 文書 304 内のリンクを選択された場合 (S109) は、翻訳ナビゲータ内で現在のモードが「翻訳モード」か「オリジナルモード」かどうかを判別する (S110)。この判別は翻訳ナビゲータ内部データのフラグでもかまわないし、図 10 の翻訳ナビゲータのインターフェースの状態による判別でもかまわない。

【0088】S110 での判別で「翻訳モード」と判別された場合、選択された URL を翻訳サーバに送信し (S111)、翻訳サーバでの翻訳処理 (S107) に戻る。

【0089】S110 での判別で「オリジナルモード」と判別された場合、選択された URL の WWW サーバ 100 へ直接アクセスし、HTML 文書をダウンロードする (S112)。これは通常のブラウジングの動作と同様のものである。

【0090】翻訳ナビゲータ起動中「翻訳モード」でブラウジング中に、図 10 の「オリジナルモード」ボタンが押下された場合、翻訳ナビゲータは HTML 文書表示部 203 に表示中の HTML 文書の URL を取得し、選択された URL の WWW サーバへ直接アクセスし、HTML 文書をダウンロードする (S112)。これは通常のブラウジングの動作と同様のものである。

【0091】翻訳ナビゲータ起動中「オリジナルモード」でブラウジング中に、図 10 の「翻訳モード」ボタンが押下された場合、翻訳ナビゲータは HTML 文書表示部 203 に表示中の HTML 文書の URL を取得し、S111 と同様に翻訳サーバへ送信する。

【0092】〔実施形態 2〕次に、本発明の他の実施形態についても図 1 を用いて説明する。

【0093】まず、本発明においては、オリジナル HTML

ML 文書 302 と翻訳・リンク補完済み HTML 文書 304 を翻訳サーバ内に保持していることから、これら 2 つを同時にブラウザ上に表示させることができる。図 11 に示すように翻訳ナビゲータ、すなわち翻訳ナビゲータ起動用 HTML 文書 301 に「同時表示」を実行させるためのインターフェース 53 を追加し、ブラウザ 200 の HTML 表示部 203 を図 12 のようにオリジナル HTML 文書表示部 62 と翻訳済み HTML 文書表示部 63 の 2 つの表示部に分割することで同時表示は可能である。

【0094】また「同時表示」の状態のままオリジナル HTML 文書表示部 62 において通常のブラウジングを行うのと同時に、翻訳サーバ 300 において翻訳処理を行った文書を翻訳済み HTML 文書表示部 63 に表示させることも可能である。

【0095】次に、本発明においては、翻訳済み HTML 文書 303 を保持していることから、翻訳 HTML 文書の作成に利用することも可能である。例えば、ある言語 1 の HTML 文書群があり、この文書群を言語 2 へ翻訳した文書群を作成する場合、言語 1 の HTML 文書群を全てブラウジングすることで、言語 2 に翻訳された HTML 文書群が翻訳サーバ上に作成されることとなる。

【0096】さらに、本発明においては、キャッシュデータ管理装置 305 と差分抽出装置 308 により、最新のオリジナル HTML 文書とキャッシュ済みオリジナル HTML 文書の差分を抽出することが可能であり、これを利用すれば、翻訳用に限らず、インターネット上の HTML 文書のどの部分が更新されたのかをブラウザ上で確認することが可能となる。例えば差分抽出装置 08 が抽出した差分情報を、最新の HTML 文書とは別の HTML 表示部に表示させたり、差分情報を元に最新の HTML 文書に前回からの変更情報を表示色を変えて表示したりすることが可能である。

【0097】また、本発明においては、HTML リンク補完装置 311 により HTML 文書に含まれるリンクの URL を取得しており、このリンクはブラウザ上に表示される HTML 文書の次に指定される可能性の高い URL である。この URL に対して翻訳サーバ 300 のプロキシアクセス部 304 とキャッシュデータ管理装置 305 が先読みを行いオリジナル HTML 文書 302、翻訳済み HTML 文書 303、翻訳・リンク補完済み HTML 文書 304 をキャッシュとして保存するような先読みキャッシュ機能の追加が可能である。先読みキャッシュ機能を行う際重要となるのは、先読みされた HTML 文書に含まれるリンクの URL に対しても先読みキャッシュを行うかどうかであるが、これはブラウザ上の操作によるアクセスからの先読み回数を翻訳サーバにおいて設定しておくことで対応できる。

【0098】また、翻訳済み HTML 文書がブラウザで正常に表示できるかの保証をする。つまり、ブラウザの

10

20

30

40

50

コンピュータに翻訳後の言語のフォントが存在するかどうかである。例えば日本語フォントのない日本語が表示できないコンピュータでブラウザを起動し、翻訳サーバにアクセスして日本語に翻訳されたHTML文書を表示させても日本語を表示できないのであれば無意味である。このような状況を解決するためには、「フォントの動的ダウンロード」「翻訳済み文書の画像生成」の機能を翻訳サーバ側に追加する必要がある。この機能を追加した上で、複数言語から複数言語への機械翻訳装置と、複数言語の辞書データベースを追加すれば、図13に示すような「言語選択」のインターフェース73を設ける事が可能となり、ブラウザを起動するクライアントコンピュータの環境に依存しない翻訳システムを実現することができる。

【0099】さらに、本発明に関わるものではないが、特開平10-134052号公報に記載されているURLと翻訳環境（例えば、コンピュータ技術辞書、スポーツ用語辞書、音楽用語辞書などのジャンル）との対応付けをキャッシュデータ管理装置305に追加すれば、図14に示すような辞書選択のインターフェース83を追加することが可能である。

【0100】

【発明の効果】以上説明したように、本発明によれば、以下の効果が奏される。

【0101】第1の効果は、ブラウザを利用するクライアントコンピュータの1つ1つに翻訳エンジンを有するプログラムと、そのプログラムが利用する辞書データベースを保持しておく必要がないということである。この効果は、この翻訳サーバさえ存在すればブラウザの起動できる全てのコンピュータで機械翻訳機能を実行できることに等しい。

【0102】その理由は、機械翻訳処理を行うのは翻訳サーバであり、辞書データベースを保持しておくのも翻訳サーバであるからである。また、翻訳したい文書をインターネット上で公開し、翻訳ナビゲータを使用してその文書を表示させることで、翻訳済み文書を取得することが可能であるからである。

【0103】第2の効果は、機械翻訳処理を翻訳サーバにおくことで精度の高い翻訳済みHTML文書をブラウザ上に表示できることにある。

【0104】その理由は、翻訳サーバを高性能なコンピュータとし、さらに高性能な機械翻訳装置と大規模な辞書データベースを翻訳サーバに持たせることで、各クライアントコンピュータにインストールする翻訳ソフトの翻訳処理に比べてその翻訳精度が大幅に向上するためである。

【0105】第3の効果は、翻訳済みHTML文書を高速にブラウザ上に表示できることである。

【0106】その理由は、まず第2の効果であげたことによるもので、翻訳サーバのコンピュータ性能と機械翻

訳装置の性能によるものである。また、既に1度翻訳されたHTML文書が存在する場合はさらに高速に翻訳済みHTML文書を表示させることが可能である。既に翻訳されたHTML文書に更新がかかっていなかった場合は全く翻訳処理は必用なくなり、翻訳済みHTML文書が更新されていた場合でも、更新部分のみを翻訳する差分翻訳処理により高速化されるためである。

【0107】第4の効果は、ブラウザ上で翻訳モードとオリジナルモードを切り替えが可能であるということである。

【0108】その理由は、ハイパーテキストリンクをたどって様々な文書を閲覧するブラウジングの操作を阻害しないことにある。翻訳モードでのブラウジングは全て翻訳済み文書が表示され、オリジナルモードでのブラウジングは通常のオリジナル文書が表示され、これらの切り替えは必用な時にブラウジングを行っているユーザーが切り替えるべき操作であるからである。

【図面の簡単な説明】

【図1】本発明の実施形態による翻訳機能付きブラウジング・システムの構成を示すブロック図である。

【図2】本発明の実施形態1による翻訳機能付きプロキシキャッシュサーバの動作を説明するためのフローチャートである。

【図3】HTML文書の一例である。

【図4】本発明の実施形態による翻訳機能付きプロキシキャッシュサーバによる翻訳処理の各段階での文書を説明するための図である。

【図5】本発明の実施形態による翻訳機能付きプロキシキャッシュサーバの入力する原（翻訳前）HTML文書と翻訳済みHTML文書を示す図である。

【図6】図1に示す差分抽出装置308によって抽出された差分情報の一例を示す図である。

【図7】絶対指定のURLと相対指定のURLを示す図である。

【図8】本発明の実施形態による「翻訳モード」「オリジナルモード」の切り替えが可能のようにURLが絶対指定に変換されたタグの一例を示す図である。

【図9】本発明の実施形態によるブラウザの動作を説明するためのフローチャートである。

【図10】本発明の実施形態1によるブラウザのHTML文書表示部に表示される翻訳ナビゲータ部の画面イメージの一例を示す図である。

【図11】本発明の実施形態2によるブラウザのHTML文書表示部に表示される翻訳ナビゲータ部の画面イメージの一例を示す図である。

【図12】本発明の実施形態2によるブラウザのHTML文書表示部に表示される画面イメージの一例を示す図である。

【図13】本発明の実施形態2によるブラウザのHTML文書表示部に表示される翻訳ナビゲータ部の画面イメ

10

20

30

40

50

ージの別の一例を示す図である。

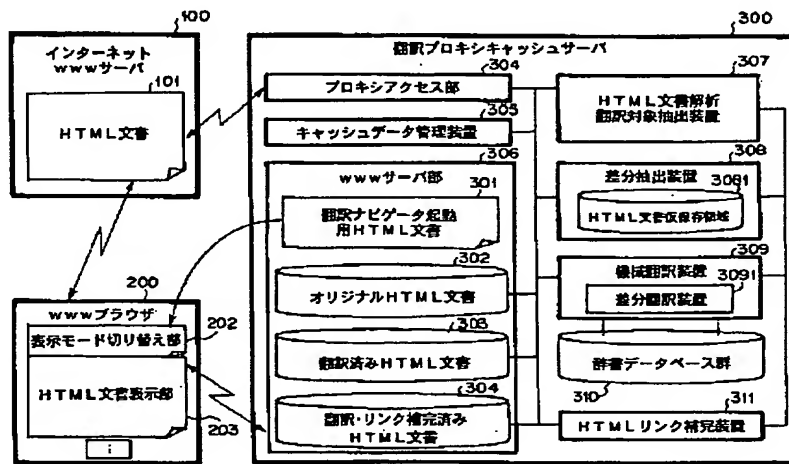
【図14】本発明の実施形態2によるブラウザのHTML文書表示部に表示される翻訳ナビゲータ部の画面イメージの更に別の一例を示す図である。

【符号の説明】

100 インターネットWWWサーバ
101 HTML文書
200 WWWブラウザ
202 表示モード切り替え部
203 HTML文書表示部
300 翻訳プロキシキャッシュサーバ

301 翻訳ナビゲータ起動用HTML文書
304 プロキシアクセス部
305 キャッシュデータ管理装置
306 WWWサーバ部
307 HTML文書解析翻訳対象抽出装置
308 差分抽出装置
309 機械翻訳装置
310 辞書データベース群
3081 HTML文書仮保存領域
3091 差分翻訳装置
311 HTMLリンク補完装置

【図1】



【図3】

HTML 文書

This is a link to Test Page .

ブラウザ上での表示

This is a link to Test Page.

【図6】

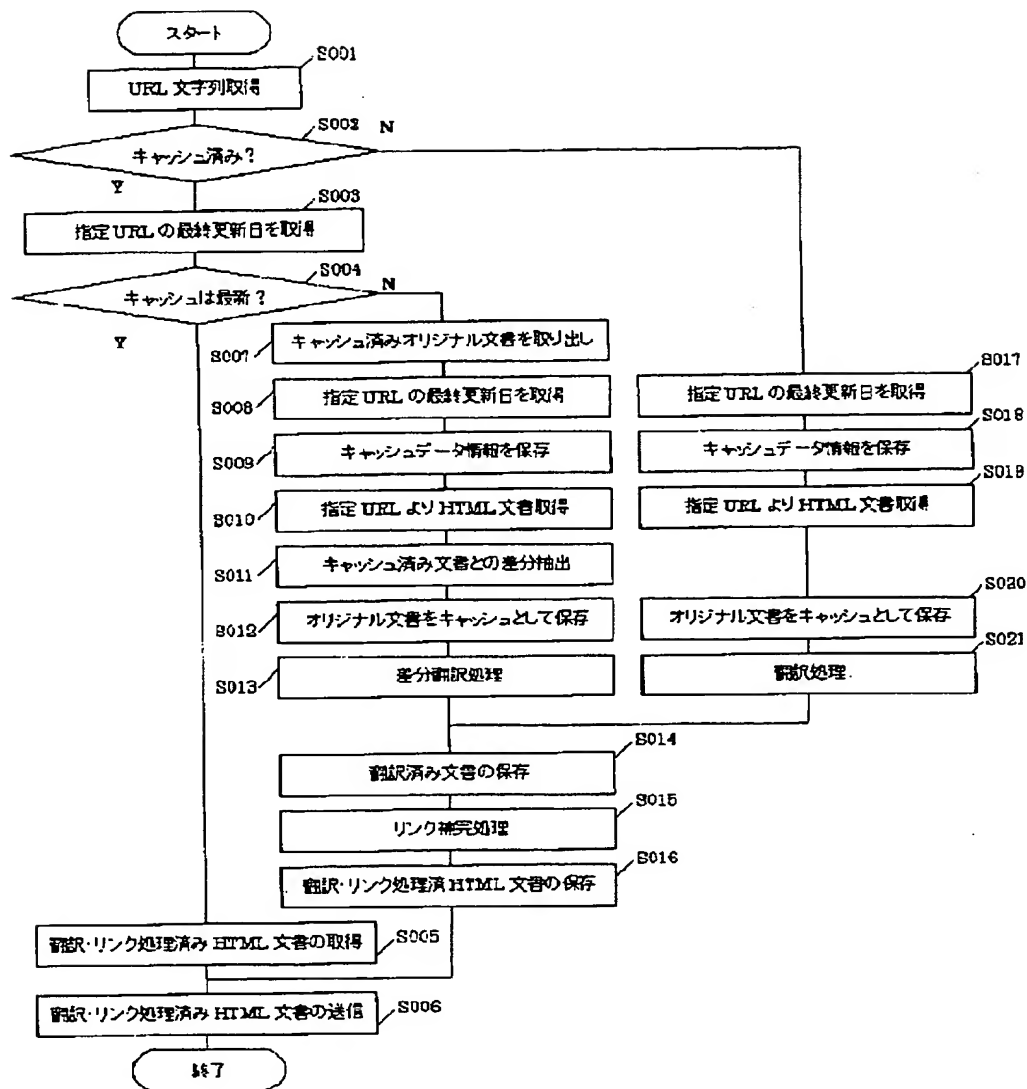
12<12

<My cat is big.

—

> My cat is sm all.

【図 2】



【図 4】

文1. オリジナル HTML 文書

 My dog is big.

文2. 文章中のタグの除去

 My dog is big.

文3. 文章の翻訳

 私の犬は大きい。

文4. 翻訳済み文書

 私の犬は大きい。

【図5】

8	<P>
9	My dog is big.
10	My dog is 10 years old.
11	<P>
12	My cat is small.
13	My dog is 1 years old.
14	

8	<P>
9	←文書構成 1 行 9 桁 0→
10	私の犬は大きい
11	私の犬は10歳です。
12	<P>
13	←文書構成 2 行 12 桁 0→
14	私の猫は小さい
15	私の猫は1歳です。

【図7】

リンクタグ 1.

 テスト

リンクタグ 2.

 同一ディレクトリ

リンクタグ 3.

 2階層上位ディレクトリ

リンクタグ 4.

 最上位からのパス

【図 8】

この翻訳済み HTML 文書の URL が `http://www.???co.jp/AAA/BBB/sample.html` の場合

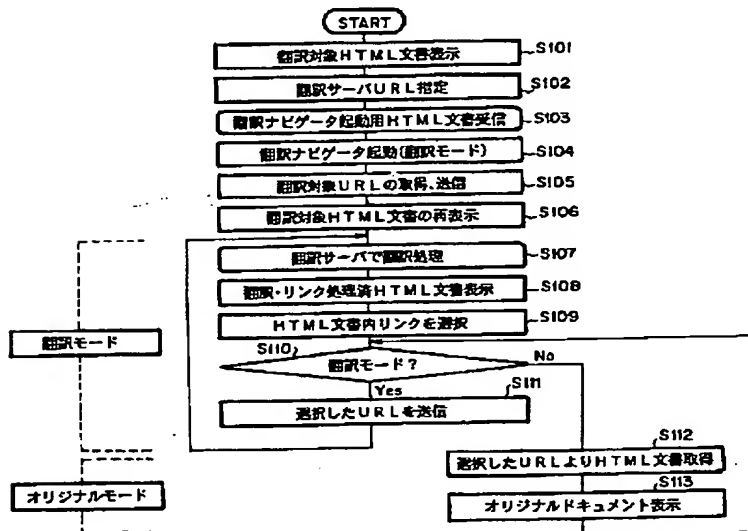
リンクタグ 1.
` テスト `

リンクタグ 2.
` 同一ディレクトリ `

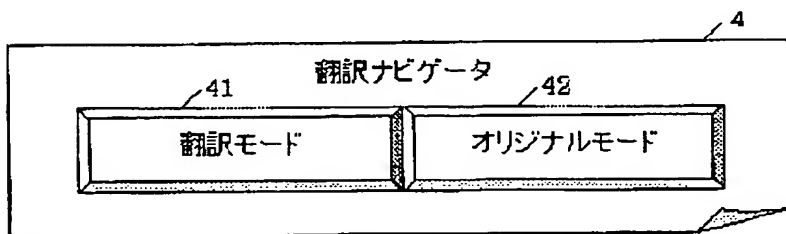
リンクタグ 3.
` 2階層上位ディレクトリ `

リンクタグ 4.
` 最上位からのパス `

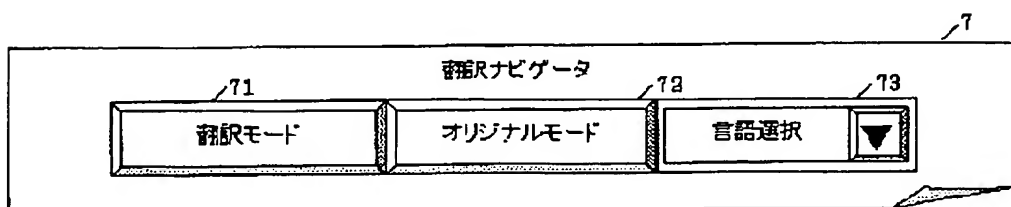
【図 9】



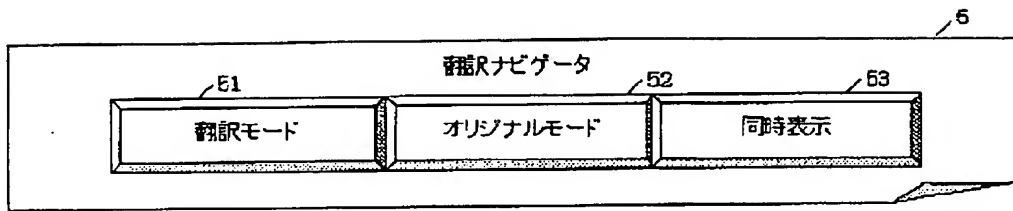
【図 10】



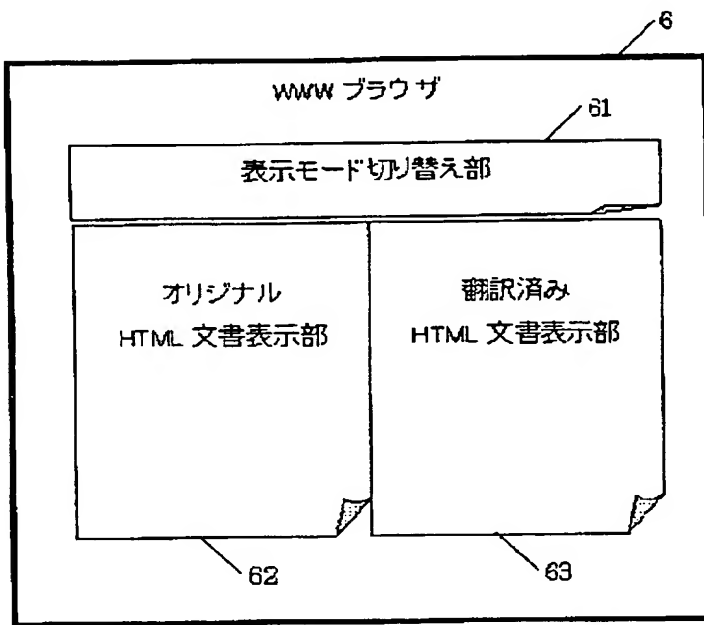
【図 13】



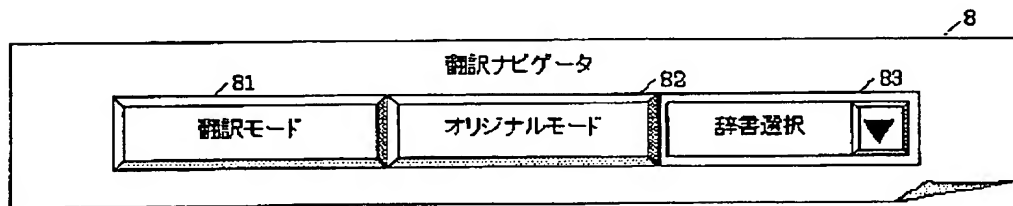
【図 11】



【図 12】



【図 14】



**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.